N89-19879

Machine Vision for Space Telerobotics and Planetary Rovers

Brian H. Wilcox
Supervisor- Robotics Systems Research,
Implementation, and Integration Group
Jet Propulsion Laboratory
California Institute of Technology
4800 Oak Grove Drive
Pasadena, California 91109
(818) 354-4625

ABSTRACT

Machine vision allows a non-contact means of determining the three-dimensional shape of objects in the environment, enabling the control of contact forces when manipulation by a telerobot or traversal by a vehicle is desired. Telerobotic manipulation in Earth orbit requires a system that can recognize known objects in spite of harsh lighting conditions and highly specular or absorptive surfaces. Planetary surface traversal requires a system that can recognize the surface shape and properties of an unknown and arbitrary terrain. Research at JPL on these two rather disparate types of vision systems is described.

INTRODUCTION

The JPL Robotics Laboratory has been conducting sensing and perception research since the mid 1970's, when a task was undertaken to develop a breadboard Mars rover which could navigate autonomously over unknown terrain. At that time, and continuing to the present, the principal sensor modality addressed was machine vision. This arises from the fact that it is essential, both in planetary rover and orbital tasks, to sense the environment prior to actual physical contact so that contact forces can be controlled. The available non-contact sensing techniques are limited to those based on electromagnetic radiation and those based on sound. Obviously sound is not useful in vacuum and of limited use in extremely rarified atmospheres. Electromagnetic sensing can be of an active type, emitting radiation and sensing the reflection, or passive, relying on ambient radiation. Active sensing systems can give direct information such as object range, but often consume excessive power and involve mechanical scanning devices which are potentially unreliable. Thus passive electromagnetic sensing is an attractive means of accomplishing the non-contact sensing function. The only wavelengths for which large amounts of ambient radiation exist in space are those emitted by the Sun, i.e. visible light and near IR. Sensors for these wavelengths are readily available with very good spatial and temporal resolution and accuracy in the form of solid-state video cameras. This has the further advantage that the human operator can easily comprehend the raw data from these sensors using a video display.

More recently, machine vision research at JPL has been extended to applications for near-Earth orbit. A useful space telerobot for on-orbit assembly, maintenance, and repair tasks must have a sensing and perception subsystem which can provide the locations, orientations, and velocities of all relevant objects in the work environment. Examples of the potential uses of such technology are robotic systems for capturing satellites which have arbitrary and unknown motion, and robotic systems for construction in space.

VISION FOR SPACE TELEROBOTICS

The sensing and perception subsystem of the Telerobot Testbed at JPL is designed to acquire and track objects moving and rotating in space, and to verify the locations of fasteners, handles, and other objects to be contacted or avoided during the space task. This system uses an array of three fixed 'wing' cameras and two cameras mounted as a stereo pair on a robot arm, which permits them to be aimed at specific objects of interest from good view angles. Processing is performed by custom image-processing hardware and a general purpose computer for high-level functions. The image-processing hardware, originally IMFEX (for Image Feature Extractor) and being upgraded to PIFEX (for Programmable Image Feature Extractor) is capable of large numbers of operations on images and on image-like arrays of data. Acquisition utilizes image locations and velocities of features extracted by the feature extractor to determine the 3-dimensional position, orientation, velocity and angular velocity of an object.

PIFEX has been described in more detail elsewhere [1][2].

The organization of the acquisition and tracking system is shown in Figure 1. The Feature Tracker detects features in the images from each camera, tracks them as they move over time, smooths their two-dimensional positions, and differentiates the positions to obtain their two-dimensional velocities in the image plane. When enough features are being tracked, the Motion Stereo module uses the information from all of the cameras for some particular time to compute the partial three-dimensional information. The Stereo Matcher refines this information and computes estimates of the scale factor and bias. It uses a general matching process based on a probabilistic search. In this process, features from one camera are matched one at a time to features from another camera in order to build a search tree. For each combination of trial matches, a least-squares adjustment is done for the scale factor and bias that produces the best agreement of the matched features. The Model Matcher matches the three-dimensional feature positions (and any other feature information available) to those of the object model in order to determine the three-dimensional position and orientation of the object [3]. Meanwhile, the Feature Tracker, running concurrently with the other modules, still has been tracking the features (those that have remained visible). The latest positions of these features, together with the information from the model matcher that indicates which object features they match, are used by the Tracking Initializer to update the object position and orientation to the time of this most recent data. The position, orientation, velocity, angular velocity, and their covariance matrix from the Tracking Initializer are used as initial conditions in the Object Tracker. It rapidly and

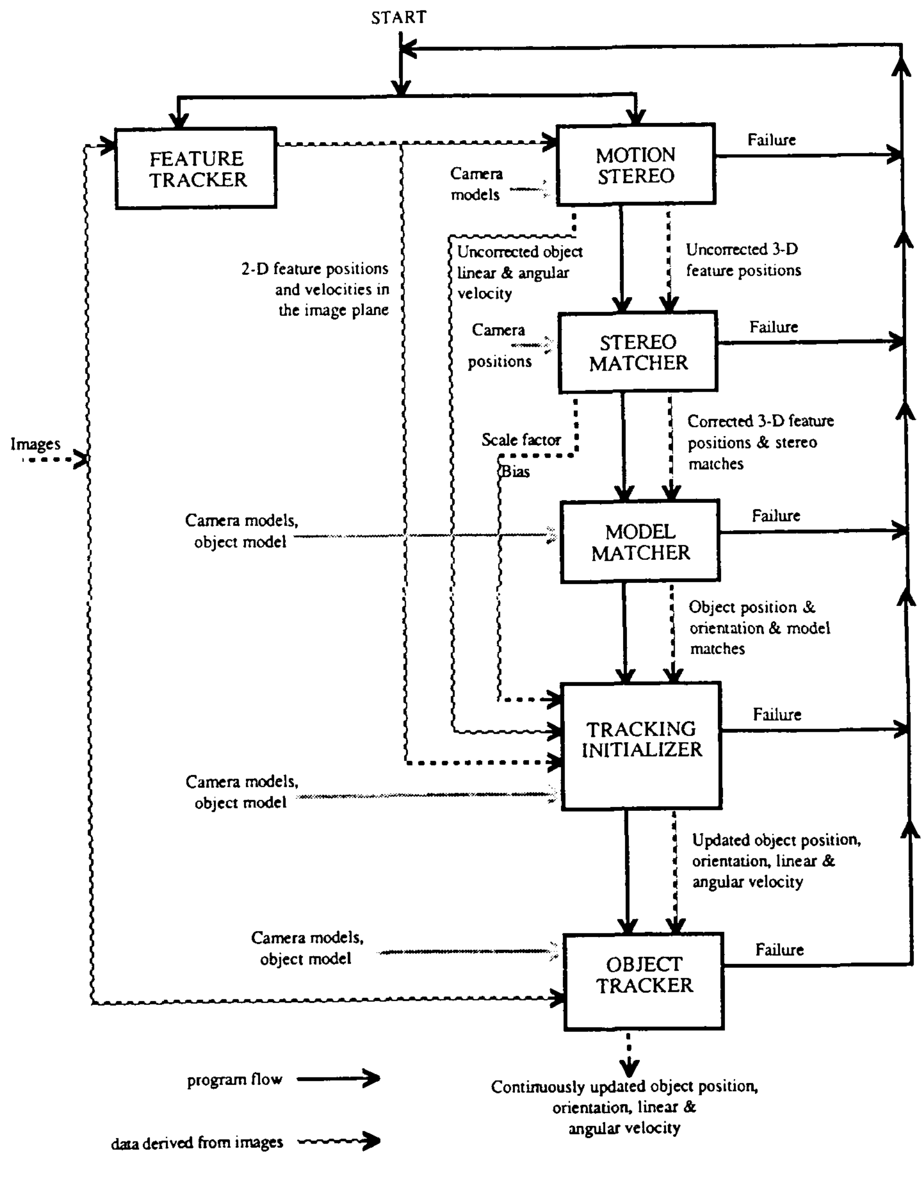

Figure 1. Acquisition and Tracking Architecture

accurately updates this information. Currently, the features that it looks for in the images are the object edges. Using edges produces more complete information than using vertices. Edges can be used easily here, because the one-dimensional information from edge elements suffices once the approximate object position and orientation are known [4].

More complete descriptions of the acquisition and tracking system have been published elsewhere [5].

VISUAL NAVIGATION AND HAZARD AVOIDANCE

Because of the long signal time to Mars (anywhere from 6 minutes to 45 minutes for a round trip at the speed of light), it is impractical to have a rover on Mars (the nearest candidate for a planetary rover) that is teleoperated from Earth (that is, one in which every individual movement would be controlled by a human being). Therefore, some autonomy on the rover is needed. On the other hand, a highly autonomous rover (which could travel safely over long distances for many days in unfamiliar territory without guidance from Earth and obtain samples on its own) is beyond the present state of the art of artificial intelligence, and thus can be ruled out for a rover launched before the year 2000.

Semiautonomous navigation is an intermediate between these two extremes. In this technique, local paths are planned autonomously using images obtained on the vehicle, but they are guided by global routes planned less frequently by human beings using a topographic map, which is obtained from images captured from a satellite orbiting Mars. The orbiter could be a precursor mission which would map a large area of Mars in advance, or it could be part of the same mission and map areas only as they are needed. As commanded from Earth, the orbiter would take a stereo pair of pictures (by taking the two pictures at different points in the orbit) of an area to be traversed (if this area is not already mapped). These pictures might have a resolution of about one meter, although poorer resolution could be used. The pictures are sent to Earth, where they are used by a human operator (perhaps with computer assistance) to designate an approximate path for the vehicle to follow, designed to avoid large obstacles, dangerous areas, and dead-end paths. This path and a topographic map for the surrounding area are sent from Earth to the rover. This process repeats as needed, perhaps once for each traverse between sites where experiments are to be done, or perhaps once per day or so on long traverses.

The sequence of operations taking place on Mars is as follows. The rover views the local scene and, by means discussed below, computes a local topographic map. This map is matched to the local portion of the global map sent from Earth, as constrained by knowledge of the rover's current position from other navigation devices or previous positions, in order to determine the accurate rover position and to register the local map to the global map. The local map (from the rover's sensors) and the global map (from the Earth) are then combined to form a revised map that has high resolution in the vicinity of the rover. This map is analyzed by computation on the rover to determine the safe areas over which to drive. A new path then is computed, revising the approximate path sent from Earth, since with the local high resolution map small obstacles can be seen which might have been missed in the low-resolution pictures used on Earth. Using the revised path, the rover then drives ahead a short distance (perhaps ten meters), and the process repeats.

With the computing power that it will be practical to put on a Mars rover in the 1990's, the computations needed to process a stereo pair of images and perform the other calculations needed may require roughly 60 seconds. If these are needed every 10 meters and it takes the rover 30 seconds to drive 10 meters, the resulting average rate of travel is 10 meters every 90 seconds, which is 11 cm/sec or 10 km/day. If a ten-kilometer path is designated from Earth each time, only one communication per day is needed, and the rover could continue to drive all night, using strobe lights for illumination. On the other hand, the method is more reliable than autonomous operation, because of the human guidance and the overview that the orbital data provides.

There are several types of computations that need to be done on the rover (or on a Mars orbiter in constant communication with the rover). These include the computation of a depth map, the computation of a topographic map, the matching of this map to the global data base and merging with it, analyzing the traversability of the area, planning a path, and monitoring the execution of the path. Some of these now will be discussed in more detail.

The first step in the processing on the rover is the production of the depth map (the distances to densely packed points over the field of view of the sensing device). One way of obtaining this is with a scanning laser range finder, which produces the depth map directly. Another way is to use two or more cameras for stereo vision. By the usual stereo process of matching and triangulation the depth map is computed. Other computer vision techniques, such as shape from shading and texture analysis, can aid in this process. Each approach has advantages and disadvantages. Stereo vision usually is more accurate at close ranges but less accurate at long ranges than laser range finders. On the other hand, laser range finders are limited in the maximum range at which they are effective. Laser range finders tend to make fewer errors than stereo vision, but each can fail to produce results under different conditions. Stereo vision and other computer vision techniques are attractive since stereo cameras will be available for scientific sample designation or core locating purposes in any event. Also, solid state cameras are small, have no moving parts, and take less power than laser scanners. Stereo matching is a computation-intensive process which is benefiting greatly from recent advances in microelectronic fabrication, while scanning mirrors remain prone to mechanical problems. Most likely, a rover should use both a scanning laser range finder and stereo cameras, to produce the best results by combining their measurements and to provide reliability in case of failure.

Recent research efforts at JPL have studied various types of stereo matching. First, we have explored the use of sophisticated statistical algorithms to reduce the error rates of two-camera stereo. Second, we are exploring the use of additional cameras and multiple cross-correlation to reduce the error rate (and possibly reduce the computational load, due to the need to match smaller patches of the images in order to achieve a given level of reliability). By using a linear array of multiple cameras which are mounted parallel so that matching between cameras is along corresponding scan lines, special hardware could be built to implement matching algorithms at frame rate. Lastly, techniques of multiresolution pyramid decomposition have been employed (where a succession of low-pass and band-pass images are produced from the original image). In this multiresolution technique, objects are matched at low resolution, and then these matches are used to guide the search at the higher resolutions. All of these techniques show promise in producing reliable depth maps and they can be combined in various ways. All of these techniques use small image patches, which are correlated between images. These are called area-based techniques, and differ from the feature-based techniques (using edges or vertices) used in vision for space telerobots. The primary difference results from the fact that spacecraft (and man-made objects in general) have a relatively small number of well-defined visual features, while natural terrain has a very large number of edges and vertices, and so is not compactly represented by simple feature extraction.

Once acquired, the depth map is transformed into an elevation map (altitudes for densely but unequally spaced horizontal positions). An important issue is whether to keep the data in the iconic form of the elevation map, in which case the topographic map sent from Earth also would be in this form, or to reduce the data to a more symbolic form.

In the iconic case, the elevation map from one view is merged with the elevation map in the data base by a process of correlation and averaging, which also produces the best estimate of vehicle position as that which produces the best correlation. (Information other than elevations, such as reflectance, could be used also.) However, this computation is more complicated than ordinary correlation because the points are not equally spaced, there may be significant uncertainties in their horizontal positions, and there may be occasional mistakes in the stereo data.

In the symbolic case, some description of objects in the scene would be extracted, for example ellipsoidal approximations of rocks [6] together with descriptions of ground slope. The same type of description would be developed from the orbital images on Earth and sent to the rover, and the matching and merging process would use these symbolic descriptions [7]. Here the techniques of vision for natural terrains begin to converge with the techniques for space telerobotics, in that the symbolic representations can be viewed as a form of feature extraction. However, these features are much more complex (e.g. 'rock' or 'crater') than those used for man-made objects ('edges' or 'vertices'). Thus low-level processing and special hardware are not generally applicable to this type of feature extraction.

With either kind of description, the local data are merged with the global data base to produce an updated data base. In some cases, each new view would be merged immediately with the global data base. However, in some cases, the matching process may not be able to correlate accurately with the global data base because of the lack of prominent features, but there may be enough smaller features to correlate with the high-resolution views seen previously from nearby locations. Therefore, a local data base could be built up by merging several local views. Then when suffiently prominent features are encountered to match well to the global data base, the local data base would be merged with it. In general, there could be a hierarchy of data bases produced in this manner.

Traversability can be determined by analyzing the data base to determine the slope and roughness of the ground at each horizontal position. This can be done by local least-square fits of planar or other surfaces and analysis of the residuals. A way of doing this for the iconic representation will be tried in the current JPL project. (If the data base is in symbolic form, the information may already be there in the form needed.)

More complete descriptions of the Mars Rover local navigation and hazard avoidance process have been published elsewhere [8].

CONCLUSIONS

Machine vision will be an important element of both space telerobots and planetary rovers. Generally, the vision systems of space telerobots will use feature extractors to generate a reduced representation of the scene, and feature-based matching in multiple cameras to generate 3-D representations. Planetary rovers, on the other hand, will use area-based scene matchers (as well as active techniqes such as laser scanning, which are less useful on orbital tasks due to the highly absorptive and specular surfaces employed) to determine the 3-D geometry of the scene. Once the 3-D geometry is known, space telerobots will generally try to recognize known objects or generic classes of items, such as fasteners. Planetary rovers may never need to assign symbolic names to objects in a scene-- a map of elevation, slope, roughness, estimated surface friction and load-bearing strength may be all that is needed. It is only a more sophisticated rover which needs to reason about landslides, unstable rocks, or 'box canyons' that may need any symbolic representation at all. Thus it may be that the two types of vision system remain quite distinct in their development, hardware, and implementation for many years to come.

ACKNOWLEDGMENT

The machine vision systems described in this paper have been created by the staff of the Robotics Systems Research, Implementation, and Integration group at JPL. The algorithms described are almost entirely the creation of staff member Dr. Donald B. Gennery. The research described in this publication was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

REFERENCES

1. D. B. Gennery and B. Wilcox, "PIFEX: An Advanced Programmable Pipelined-Image Processor," Jet Propulsion Laboratory Publication 84-97, 1984.
2. D. B. Gennery and B. Wilcox, "A Pipelined Processor for Low-Level Vision," *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition* San Francisco, CA, June 1985, pp. 608-613.
3. D. B. Gennery, "A Feature-Based Scene Matcher," *Proceedings of the Seventh International Joint Conference on Artificial Intelligence,* Vancouver, August 1981, pp. 667-673.
4. D. B. Gennery, "Tracking Known Three-Dimensional Objects," *Proceedings of the AAAI Second National Conference on Artificial Intelligence* , Pittsburgh PA, August 1982, pp. 13-17.
5. D. B. Gennery, "Stereo Vision for the Acquisition and Tracking of Moving Three-Dimensional Object," in *Techniques for 3-D Machine Perception* (A. Rosenfeld, ed.), North Holland, 1985.
6. D. B. Gennery, 'Object Detection and Measurement Using Stereo Vision,' *Proc. Sixth International Joint Conference on Artificial Intelligence,* Tokyo, Japan, pp. 320-327 (1979).
7. D. B. Gennery, 'A Feature-Based Scene Matcher,' *Proc. Seventh International Joint Conference on Artificial Intelligence*, Vancouver, British Columbia, pp. 667-673 (1981).
8. B. Wilcox and D. Gennery, 'A Mars Rover for the 1990's', *Journal of the British Interplanetary Society*, Vol. 40, No. 10, pp. 483-488, Oct 1987.

C - 6